



A multicriteria decision support tool for modelling bank credit ratings

Chrysovalantis Gaganis¹ · Panagiota Papadimitri² · Menelaos Tasiou^{2,3} 

© The Author(s) 2020

Abstract

We introduce an application of the SMAA-Fuzzy-FlowSort approach to the case of modelling bank credit ratings. Its stochastic nature allows for imprecisions and uncertainty that naturally surround a decision-making exercise to be embedded into the proposed framework, whilst its output complements the ordinal nature of a crisp classification with cardinal information that shows the degree of membership to each rating category. Combined with the SMAA variant of GAIA that offers a visual of a bank's judgmental analysis, both recent approaches provide a holistic multicriteria decision support tool in the hands of a credit analyst and enable a rich inferential procedure to be conducted. To illustrate the assets of this framework, we provide a case study evaluating the credit risk of 55 EU banks according to their financial fundamentals.

Keywords Bank credit ratings · MCDA · SMAA · FlowSort · GAIA

JEL Classification G21 · G24 · C38 · C53

1 Introduction

Over the past three decades the credit quality of financial institutions became a central theme in the agenda of governments, investors and depositors. The Basel II accord uses the credit ratings to provide a more refined measure of banks' credit risk exposure (Pasiouras

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10479-020-03516-9>) contains supplementary material, which is available to authorized users.

✉ Menelaos Tasiou
menelaos.tasiou@port.ac.uk

Chrysovalantis Gaganis
c.gaganis@uoc.gr

Panagiota Papadimitri
panagiota.papadimitri@port.ac.uk

¹ Department of Economics, University of Crete, Rethymno, Greece

² Portsmouth Business School, University of Portsmouth, Portsmouth, UK

³ Centre for Operations Research and Logistics, University of Portsmouth, Portsmouth, UK

et al. 2007). Their importance and use is central also in the Basel III framework. In particular, credit ratings are important for the risk assessment of assets, as well as assessments of various exposures (BIS 2011). Understandably, the prominence of leading major rating agencies such as Standard and Poor's Ratings Services (S&P's), Moody's Investors Service (Moody's) and Fitch Investors Service (Fitch)—which play a key role in the financial system using model risk measures to produce credit risk ratings—is signified.

A large number of statistical and machine learning techniques have been used for the examination of the credit ratings of banks (Poon et al. 1999; Poon and Firth 2005; Pasiouras et al. 2006, 2007; Demircuc-Kunt et al. 2006; Gaganis et al. 2006). Multicriteria Decision Aiding (MCDA) methods (Greco et al. 2016) boast a handful noteworthy characteristics, making them especially useful when it comes to this domain (Doupou and Figueira 2019). Generally speaking, MCDA methods have been applied to several areas of finance (see e.g. Doupou and Zopounidis 2014, for an in-depth overview) with a great success and trending interest (see e.g. Zopounidis et al. 2015, for a bibliographic survey), eventually proving to be great assets in supporting decision-makers' (DMs) financial decisions. In the strand of credit rating analysis that this study is addressed at in particular, MCDA models have been used in a variety of assessment and predictive frameworks, making use of value functions (Doupou and Pasiouras 2005; Gavalas and Syriopoulos 2015), rough sets theory (Capotorti and Barbanera 2012), goal programming (García et al. 2013) and outranking techniques (Angilella and Mazzù 2015; Doupou and Figueira 2019).

This study complements the latter modelling framework by employing the recently introduced SMAA-Fuzzy-FlowSort (hereafter referred to as 'SMAA-FFS'; Pelissari et al. 2019a), the first Stochastic Multiobjective Acceptability Analysis (SMAA) (Lahdelma et al. 1998; Lahdelma and Salminen 2001) variant of PROMETHEE-based sorting¹ methods. The framework of SMAA is used as a means to deal with imperfections and uncertainty in real world applications, and it has a wide gamut of applicability that extends and crosses several disciplines (see Pelissari et al. 2019b for a recent survey). Of course, turning to the case of our interest in particular, an effective credit risk assessment/modelling is well in line with the characteristics of the proposed method for a variety of reasons that we list forthwith.

First and foremost, credit rating criteria boast heterogeneous scales that are often regarded having imperfections, making the imposition of crisp decision rules on the criteria scales to describe the creditworthiness of firms rather difficult (Doupou and Figueira 2019). By extension, dealing with weight elicitation in real life is often difficult due to the inherent complexity in defining a single and well-defined weight vector that denotes the importance of criteria (Greco et al. 2019); a challenging task, which still remains one of the most relevant questions in the MCDA domain (Belton and Stewart 2002; Greco et al. 2016). Moreover, being based on the SMAA framework, SMAA-FFS does not only model uncertainty as to the previous two aspects but further enhances the transparency and robustness of the evaluation exercise using the capacity of SMAA to take into account the whole space of feasible parameters (Lahdelma et al. 1998; Lahdelma and Salminen 2001). In fact, probabilistic outcomes such as the category acceptability indices produced within SMAA-FFS (to be detailed in Sect. 2) seem more plausible and realistic in such an analysis

¹ PROMETHEE I & II methods to rank alternatives have been earlier introduced in the SMAA environment by Corrente et al. (2014).

rather than a simple crisp classification setting based on a single and supposedly representative vector of parameters, the robustness and validation of which could be regarded as a matter of dispute. Besides, as it will be argued later on, such probabilities could be seen as a cardinal measure of risk rating that complements the ordinal nature of a crisp classification with information showing the potential membership of a bank into each rating group.

A second reason the proposed method is a match for judgmental analysis lies on the fact that, being a member of the PROMETHEE family of methods (Brans and Vincke 1985; Brans et al. 1986; Brans and De Smet 2016) provides the DM with tools such as the GAIA plane (Mareschal and Brans 1988), which further supports him through graphical representations of the alternatives' performance. Here we will make use of the SMAA variant of GAIA by Arcidiacono et al. (2018) that visualises any kind of ordinal information on the plane based on the space of feasible parameters taken into account in the evaluation phase; an act that is well in line with the ordinal nature of credit ratings.

Given the above, in this study we advocate the use of SMAA-FFS as a potential alternative to the modelling framework of credit ratings. To the best of our knowledge, the closest studies to our own in this domain are those of Angilella and Mazzù (2015) and Doumpos and Figueira (2019), making use of the ELECTRE TRI (Roy and Bouyssou 1993) on a SMAA-based environment (Tervonen et al. 2007) and ELECTRE TRI-Nc (Almeida-Dias et al. 2012) methods to construct judgmental rating models for SMEs and industrial firms respectively. Whilst still belonging to the outranking family of MCDA methods (i.e. both being ELECTRE-based methods) hence sharing many of the characteristics that we mentioned above, the alternative proposed in this paper has the following two benefits. First, most methods under the PROMETHEE family—such as the one we propose—, require fewer parameters than other MCDA methods, such as ELECTRE TRI. Put simply, the first benefit relates to a reduced cognitive stress placed onto the DM, or one parameter less to take into account in the simulation environment. Second, the use of the SMAA variant of GAIA that only exists in PROMETHEE-based methods is an attractive tool in the hands of the credit analyst that enables visualisation of the space of preferences according to which the bank of interest achieves a given ordinal outcome, as well as a tool for sensitivity analysis where elicitation of a single weight vector is possible.

Let us also point out here that an early study on the use of PROMETHEE methods for credit analysis has been conducted by Doumpos and Zopounidis (2010). The authors present a specific case study on the implementation of such an approach on Greek banks in cooperation with expert analysts from the Bank of Greece. Of course, the combination of PROMETHEE II and SMAA methods has formally been introduced later by Corrente et al. (2014), along with the output that SMAA entails (to be detailed in Sect. 2), bringing more insights regarding the probabilistic outcomes of interest to a DM. Our difference with the study of Doumpos and Zopounidis (2010) also includes, but is not limited to, the introduction of reference profiles that are implicit in the FlowSort method—instead of the use of predetermined interval thresholds on the obtained overall (normalised) net flows applied ex post—, the use of SMAA's probabilistic categorical classifications and the use of SMAA-GAIA to enhance the toolbox of an analyst with visual illustrations of the output according to the space of preferences.

Last but not least, we should hereby mention that, as in previous similar studies utilising outranking approaches for credit analysis (Angilella and Mazzù 2015; Doumpos and Figueira 2019), the proposed framework's objective is not to serve as an 'optimally'

predictive/reproduction tool of external credit ratings. Rather, the proposed framework could be used for judgmental credit risk rating in a setting of constructing internal corporate credit rating models using a set of financial fundamentals and benchmarking the results to the externally received ratings from other credit rating agencies (CRAs). Therefore, on a similar note, we extend that particular strand of literature regarding the use of MCDA models for judgmental credit ratings to test, validate and make inferences about the soundness of a bank and identify potential weaknesses to work on and improve. To illustrate how such an analysis could be conducted and be of a great help in the hands of a credit analyst, we evaluate in a simulation environment a set of 55 EU banks rated by Standard & Poor's (S&P) over a period of 6 years.

In what follows, Sect. 2 presents the methodological preliminaries of the SMAA-FFS and SMAA-GAIA methods. Section 3 contains a description and preliminary analysis of the data and Sect. 4 provides the obtained results, some additional analysis and a discussion. Finally, Sect. 5 concludes this study and provides direction for future research.

2 Preliminaries

In this section, we give an abridged overview of the SMAA-FFS and SMAA-GAIA approaches that will be referred to and used in the subsequent part of the paper. As it will become clear towards the end of this section, these methodologies are all linked together to be used as one modular tool that evaluates (SMAA-FFS) and visualises (SMAA-GAIA) the set of alternatives based on a set of criteria chosen by the DM. To conserve space, we will only refer to those features that we actually use in this paper.

2.1 The PROMETHEE methods

'PROMETHEE' refers to a set of approaches (see Brans and De Smet 2016, for a recent review) that belongs to the broader family of outranking MCDA methods (Greco et al. 2016). The PROMETHEE I & II methods (Brans and Vincke 1985; Brans et al. 1986) are quite popular in a broad range of disciplines across the academic community [see Behzadian et al. (2010) and Zopounidis et al. (2015), for two recent surveys] and were designed as ranking methods that are based on pairwise comparisons of alternatives that share several interesting features. In particular, PROMETHEE methods aggregate the preference information shared by a DM through valued preferences relations that are described forthwith.

Suppose a set of banks $A = \{a_1, \dots, a_n\}$, $i \in I = \{1, \dots, n\}$ to be evaluated based on some risk assessment criteria $G = \{g_1, \dots, g_m\}$, where $g_j : A \rightarrow \mathbb{R}$, $j \in J = \{1, \dots, m\}$. For each criterion $g_j \in G$, PROMETHEE constructs a preference function $P_j(a_i, a_{i'})$, $i' \neq i$ and $i' \in I$ that essentially represents the degree of preference of bank a_i over another bank $a_{i'}$ on criterion g_j being a non-decreasing function of $d_j(a_i, a_{i'}) = g_j(a_i) - g_j(a_{i'})$. There are six alternative functions that could be considered for each and every criterion on behalf of a DM (Brans and De Smet 2016), though we hereby choose the piecewise linear function for reasons of simplicity. This is defined as follows:

$$P_j(a_i, a_{i'}) = \begin{cases} 0 & \text{if } d_j(a_i, a_{i'}) \leq q_j \\ \frac{d_j(a_i, a_{i'}) - q_j}{p_j - q_j} & \text{if } q_j < d_j(a_i, a_{i'}) < p_j \\ 1 & \text{if } d_j(a_i, a_{i'}) \geq p_j \end{cases},$$

where q_j and p_j are the indifference and preference thresholds accordingly, as these are given by the DM for each criterion $g_j \in G$. Risk assessment criteria are attached an importance w_j , with $w_j \geq 0$ and $\sum_{j=1}^m w_j = 1$. For each pair of banks $(a_i, a_{i'}) \in A \times A$, PROMETHEE methods compute how much a bank a_i is preferred over $a_{i'}$ taking into account all criteria $g_j \in G$ as follows:

$$\pi(a_i, a_{i'}) = \sum_{j=1}^m w_j P_j(a_i, a_{i'}),$$

with values of $\pi(a_i, a_{i'})$ ranging between 0 and 1. Obviously, higher values denote higher preference and vice versa. Making these comparisons meaningful and consistent across all pairs, comparing a bank a_i with all remaining banks $a_{i'}$ is made feasible through the computation of the positive and negative flows as follows:

$$\varphi^+(a_i) = \frac{1}{n-1} \sum_{a_{i'} \in A} \pi(a_i, a_{i'}), \quad \text{and} \quad \varphi^-(a_i) = \frac{1}{n-1} \sum_{a_{i'} \in A} \pi(a_{i'}, a_i).$$

The former shows how much on average a bank a_i is dominating the remaining ones taking into account all risk-assessment criteria, whilst the latter flow shows how much it is dominated on average. Understandably, the higher the positive (φ^+) and the lower the negative (φ^-) flow, the better the performance of a bank with respect to its counterparts.

PROMETHEE I compares the two flows individually using the preference (P), indifference (I) and incomparability (R) relations to provide a partial ranking of the alternatives as follows:

$$\begin{cases} a_i P a_{i'} & \text{iff } \begin{cases} \varphi^+(a_i) > \varphi^+(a_{i'}) \text{ and } \varphi^-(a_i) < \varphi^-(a_{i'}) \text{ or} \\ \varphi^+(a_i) = \varphi^+(a_{i'}) \text{ and } \varphi^-(a_i) < \varphi^-(a_{i'}) \text{ or} \\ \varphi^+(a_i) > \varphi^+(a_{i'}) \text{ and } \varphi^-(a_i) = \varphi^-(a_{i'}) \end{cases} \\ a_i I a_{i'} & \text{iff } \varphi^+(a_i) = \varphi^+(a_{i'}) \text{ and } \varphi^-(a_i) = \varphi^-(a_{i'}) \\ a_i R a_{i'} & \text{iff } \begin{cases} \varphi^+(a_i) > \varphi^+(a_{i'}) \text{ and } \varphi^-(a_i) > \varphi^-(a_{i'}) \text{ or} \\ \varphi^+(a_i) < \varphi^+(a_{i'}) \text{ and } \varphi^-(a_i) < \varphi^-(a_{i'}) \end{cases} \end{cases}$$

When incomparabilities among alternatives arise ($a_i R a_{i'}$), the use of PROMETHEE II alleviates the issue by providing a net unipolar score that combines the two flows as follows:

$$\varphi(a_i) = \varphi^+(a_i) - \varphi^-(a_i),$$

and gives a ranking of the alternatives in a complete pre-order based on the preference and indifference relations that are now changed into:

$$\begin{cases} a_i P a_{i'} & \text{iff } \varphi(a_i) > \varphi(a_{i'}) \\ a_i I a_{i'} & \text{iff } \varphi(a_i) = \varphi(a_{i'}) \end{cases}.$$

The net flow is defined in the $[-1, 1]$ range and essentially describes how much on average a bank a_i is preferred over all remaining banks taking into account at the same time how much it is dominated as well.

2.2 The FlowSort method

FlowSort is a sorting variant of PROMETHEE developed by Nemery and Lamboray (2008) for assigning alternatives into predefined ordered categories, i.e. C_1, \dots, C_h . According to the authors, categories can be described using either limiting profiles (a lower and an upper bound on each criterion characterising the class), or centroids (commonly called ‘central profiles’), e.g. based on models such as the ones proposed by Doumpos and Zopounidis (2004) that are seen as representations of a typical alternative for each category. In this paper, we find the use of the latter more appropriate, thus we will describe this procedure accordingly. Nonetheless, for the readership interested in the use of limiting profiles, a detailed description can be found in the original study of Nemery and Lamboray (2008).

Let $R = \{r_1, \dots, r_h\}$ be the set of reference profiles characterising the h credit rating categories, $k \in K = \{1, \dots, h\}$, in which r_1 is the worst possible category’s reference profile, and r_h is the best one. One could think of the set R as a set of h virtual alternatives against which the alternatives $a_i \in A$ will be benchmarked. Understandably, the evaluation of banks is also delimited by r_1 and r_h . As the classes are ordered from the worst to the best, each reference profile is preferred to the one it precedes, i.e.: $r_1 < r_2 < \dots < r_h$. Now, for every bank a_i , let us define the set $R_i = R \cup \{a_i\}$, $i \in I = \{1, \dots, n\}$. Eventually, *positive*, *negative* and *net* flows can be computed as follows:

$$\varphi_{R_i}^+(a_i) = \frac{1}{|R_i| - 1} \sum_{a_{i'} \in R_i} \pi(a_i, a_{i'}),$$

$$\varphi_{R_i}^-(a_i) = \frac{1}{|R_i| - 1} \sum_{a_{i'} \in R_i} \pi(a_{i'}, a_i), \text{ and } \varphi_{R_i}(a_i) = \varphi_{R_i}^+(a_i) - \varphi_{R_i}^-(a_i).$$

Using the above output permits classifying a bank into a specific predefined category following the set of rules that we describe below, which is, of course, a generalisation of the ranking relations found in PROMETHEE I and II methods. That is:

$$C_{\varphi^+}(a_i) = C_h \text{ if } \frac{\varphi_{R_i}^+(r_{h-1}) + \varphi_{R_i}^+(r_h)}{2} < \varphi_{R_i}^+(a_i) \leq \frac{\varphi_{R_i}^+(r_h) + \varphi_{R_i}^+(r_{h+1})}{2} \text{ and}$$

$$C_{\varphi^-}(a_i) = C_h \text{ if } \frac{\varphi_{R_i}^-(r_{h-1}) + \varphi_{R_i}^-(r_h)}{2} \geq \varphi_{R_i}^-(a_i) > \frac{\varphi_{R_i}^-(r_h) + \varphi_{R_i}^-(r_{h+1})}{2},$$

for the PROMETHEE I comparable sorting framework, and:

$$C_{\varphi}(a_i) = C_h \text{ if } \frac{\varphi_{R_i}(r_{h-1}) + \varphi_{R_i}(r_h)}{2} \geq \varphi_{R_i}(a_i) > \frac{\varphi_{R_i}(r_h) + \varphi_{R_i}(r_{h+1})}{2},$$

for the equivalent PROMETHEE II sorting framework. The latter, again, helps to alleviate any incomparabilities, by ensuring that each bank can only be classified to one exact

category. For the subsequent part of the analysis, we will only use the latter relation, that is the PROMETHEE II sorting, i.e. C_φ .

The Fuzzy-FlowSort method (Campos et al. 2015) that is part of the overall SMAA-FFS approach (Pelissari et al. 2019a) we use in this study is skipped, as we do not intend to use any linguistic variables; hence, to conserve space, we refer the reader interested in this method to the original studies, describing in detail how to use it and/or how to embed it within a SMAA framework respectively.

2.3 SMAA

Stochastic Multiobjective Acceptability Analysis (SMAA) was initially proposed by Lahdelma et al. (1998), offering a solid solution to real-world decision-making problems surrounded by any source of uncertainty. The aim of the objective was originally to find out the most acceptable alternatives, which, in a second variant (SMAA-2; Lahdelma and Salminen 2001) was further enhanced to provide a DM with insights such as the probabilities with which rankings are achieved for every evaluated unit. These are made possible by considering probability distributions on the space of parameters that are surrounded by uncertainties, whether these regard the raw data of the alternatives being considered, or the DM's sets of preferences.

The fusion of SMAA and PROMETHEE methods was formally introduced by Corrente et al. (2014). The authors describe how to take into account all potential sources of uncertainty, though in this framework we focus on only three: *weights*, *preference* and *indifference thresholds*. That said, we leave out uncertainties arising from potential changes in the raw data used in the analysis (CAMEL framework proxies—to be detailed in Sect. 3). The above-mentioned sources of uncertainties could be modelled by considering a probability distribution f_w over the space of feasible weights W , as well as two probability distributions f_q and f_p over the space of feasible Q and P respectively. Assuming lack of constraints on these spaces, these can be defined as:

$$W = \left\{ \mathbf{w} = [w_1, \dots, w_m] : w_j \geq 0, j = 1, \dots, m, \sum_{j=1}^m w_j = 1 \right\},$$

for the weight space W , and:

$$P = \left\{ \mathbf{p} = [p_1, \dots, p_m] : \min_{a_i, a_{i'}} |d_j(a_i, a_{i'})| \leq p_j, j = 1, \dots, m \right\}, \text{ and}$$

$$Q = \left\{ \mathbf{q} = [q_1, \dots, q_m] : q_j \leq p_j, j = 1, \dots, m \right\},$$

for the spaces P and Q respectively. Of course, the relation between the latter thresholds q and p is such that their distributions should not overlap, always respecting the inequality constraint $q \leq p$. In the lack of any information from a DM, thresholds are uniformly distributed. Nonetheless, any kind of information provided by the DM could shape these

spaces by either considering different probability distributions, or any kind of constraints, such as assurance regions, linear inequalities etc..² An example of the latter will be given in the illustrative example in Sect. 4.

Turning to the fusion of SMAA and FlowSort (Pelissari et al. 2019a), the reference profiles could also be simulated within e.g. a given interval. In such case, the central profile becomes a random variable ψ with a probability distribution f_ψ and is defined in the space Ψ :

$$\Psi = \{\psi \in \mathbb{R}^{h \times m} : \psi_{kj} - \psi_{lj} \geq 0, \forall k > l, k = 1, \dots, h, l = 1, \dots, k-1, j = 1, \dots, m\}.$$

Taking into account the whole spaces W , Q , P and Ψ , in order to define the category acceptability index, a categorisation function $h = K(i, q, p, \psi, w)$ can be defined, showing the category h in which a bank a_i is assigned. Following from Pelissari et al. (2019a), the category membership function m_i^h can be defined as:

$$m_i^h(q, p, \psi, w) = \begin{cases} 1, & \text{if } K(i, q, p, \psi, w) = h \\ 0, & \text{otherwise} \end{cases}.$$

Finally, the category acceptability index C_i^h can be computed using numerical integration over the spaces of feasible parameters taken into account in the model, which hereby means:

$$C_i^h = \int_{p \in P} f_p(p) \int_{q \in Q} f_q(q) \int_{\psi \in \Psi} f_\psi(\psi) \int_{w \in W} f_w(w) m_i^h(q, p, \psi, w) dw d\psi dq dp.$$

Essentially, category acceptability indicators can be translated into probabilistic classification taking into account the feasible spaces of parameters assumed in an evaluation (Lahdelma and Salminen 2010). In the concept of bank credit rating classification, this shows the degree of membership of a bank into each class, according to the preferences taken into account in the evaluation environment. Let us also note here that, another interesting insight could be obtained by computing the typical preferences of a DM, for which a bank is probabilistically assigned at a category of interest, e.g. perhaps closer to default (feasible through the computation of central weight vectors, see e.g. Corrente et al. 2014).

Reaching a single classification based on the above probabilistic output is feasible using the category acceptability indicators. This can be done in two ways. First, as Lahdelma and Salminen (2010) point out, one approach is to use the median category, i.e. that first class by which 50% of the probability mass is accumulated. This is an indicative rule of majority and, of course as the authors also point out, this is subjective in the sense that it depends on the risk attitude of the decision maker who could decide on another fractile. In this study, we take advantage of the whole output by using holistic indicators originally introduced in SMAA-2 (Lahdelma and Salminen 2001). These are essentially combinations of the rank acceptability (hereby category acceptability) indicators using some meta-weights to convert them into a single index that portrays this information. Following from the authors (Lahdelma and Salminen p. 449):

² For a list of such feasible ways to use constraints, see e.g. the study of Allen et al. (1997). Although applied to the context of Data Envelopment Analysis, the intuition behind these constraints is the same and can easily be applied in this context.

$$c_i = \sum_{h=1}^k \tau^h C_i^h,$$

where c_i is the final (holistic) classification of bank a_i , τ^h is the h -th element of the so-called meta-weights τ , and C_i^h refers to the h -th category acceptability indicator of bank a_i . The meta-weights chosen are ordinal numbers that express the rating category, i.e. $\tau = (1, 2, \dots, k)$. Put simply, the final classification combines the probabilistic indicators with their corresponding categories to find the expected category that a bank a_i belongs to. As categories refer to discrete numbers, whereas expected categories could be regarded as random continuous variables in the $[1, k]$ range, similarly to the handling of centroids, a bank a_i is classified in category C_h as follows:

$$\begin{cases} C_1 & \text{if } 1 \leq c_i < \frac{1+2}{2} \\ \vdots & \vdots \\ C_h & \text{if } h-1 \leq c_i < \frac{2h-1}{2} \end{cases}.$$

In a nutshell, a bank a_i can be classified in the category $[c_i]$ being the maximal integer contained in c_i . This output serves as the final, overall expected classification of the bank a_i .

2.4 GAIA

GAIA (Mareschal and Brans 1988) is a visual interactive module that is usually implemented alongside the PROMETHEE methods, providing DMs with a clear view on alternatives' performance in the considered criteria. Taking into account the PROMETHEE methods framework described in Sect. 2.1, the creation of the $n \times m$ unicriterion net flow matrix U is required, which is defined as follows:

$$U = \begin{pmatrix} u_{a_1, g_1} & u_{a_1, g_2} & \cdots & u_{a_1, g_m} \\ u_{a_2, g_1} & u_{a_2, g_2} & \cdots & u_{a_2, g_m} \\ \vdots & \vdots & \cdots & \vdots \\ u_{a_n, g_1} & u_{a_n, g_2} & \cdots & u_{a_n, g_m} \end{pmatrix},$$

where $u_{a_i, g_j} = \frac{1}{n-1} \sum_{i'=1}^n [P_j(a_i, a_{i'}) - P_j(a_{i'}, a_i)]$ is essentially the net dominance of a bank a_i based on criterion g_j . Put simply, GAIA reduces the m -dimensional dominance space into a two-dimensional plane that is visually clear to the keen eye. This becomes feasible by applying Principal Component Analysis (PCA) on matrix U . In particular, consider the case that one wishes to construct a two-dimensional GAIA plane, with λ_1, λ_2 the two largest eigenvalues and e_1, e_2 the corresponding eigenvectors obtained from applying PCA on matrix U . Considering it happens that the explained variance (i.e. $\delta = \frac{\lambda_1 + \lambda_2}{\sum_{l=1}^m \lambda_l}$) is at least 60% (Brans and Mareschal 1995), the GAIA plot consists of a two-dimensional plane on which:

- Each risk assessment attribute g_j is plotted with coordinates $(e_{1(j)}, e_{2(j)})$, and a line linking it to the origin of the plane $(0,0)$.
- Each bank is plotted using its principal component scores as coordinates.

- The ‘decision stick’ is plotted using the coordinates $(\mathbf{w}^T \mathbf{e}_1, \mathbf{w}^T \mathbf{e}_2)$, where \mathbf{w} is the chosen weight vector.

Arcidiacono et al. (2018) propose a SMAA variant of the GAIA plane, which, instead of a single ‘decision stick’, plots the whole (or a constrained) space of weights accordingly, displaying ordinal information according to a certain condition. For instance, the authors describe several examples in which a portion of the feasible weight vector space is highlighted (usually with a colour in the RGB gamut) denoting the space of preferences according to which an alternative is ranked 1st, 2nd etc. Adjusting this into our case, in the SMAA variant of GAIA that we will provide, we similarly highlight the space of weight vectors taken into account (i.e. 10,000 uniformly distributed weight vectors), according to which a bank a_i is classified into a category C_h .

This gives two types of information: First and foremost, the potential linear combination of attributes according to which a bank is classified into a good (bad) category, thus providing room for improvement in a given dimension, or a combination of dimensions that aggregated still give us a mediocre classification for the bank. Second, it provides a type of visual robustness analysis, illustrating the space according to which an alternative would not change a credit rating category. For instance, consider that a bank has an internal analyst that declares a univocal vector of preferences (a single set of weights); how robust is the evaluation attained according to this vector; or, in other words, how sensitive to the analyst’s perception is the classification of the bank attained?

3 Data

This section provides a description of the data used in the empirical part that is following in the subsequent sections. Section 3.1 gives the context of the analysis, while Sect. 3.2 provides some preliminary descriptive summaries of the underlying ratings and risk assessment criteria data on a concise but informative note.

3.1 Context of the analysis

Our analysis focuses on bank credit ratings, as they play an integral role within the financial system. More importantly, the rising complexity and volume of financial transactions over the past years has emphasized even more the role of credit ratings within the financial system. Ratings can be seen as mechanisms that alleviate asymmetric information issues between borrowers and lenders (Langohr and Langohr 2010). They are of fundamental importance when it comes to banks, as they represent the basis for loan approval, pricing, monitoring and loan loss provisioning (Grunert et al. 2005). In addition, bank credit ratings play an important role for policy makers, too. In particular, in the US regulatory agencies have often authorised credit rating agencies to oversee the credit quality of banks’ portfolios (White 2010). While external ratings have been established in the 20th century, banks’ interest in internal ratings started increasing considerably during the 90 s, particularly being used for analysis and reporting and administration (Treacy and Carey 2000). Their importance is analysed in great depth in Basel Committee on Banking Supervision (2011, see Section III, Risk Coverage B) along with their need to be back-tested. There are

several studies delving into the empirical analysis of banks' internal rating systems (see e.g. Grunert et al. 2005, p. 511 for a concise list).

Since the introduction of the Basel II framework, banks are allowed to adapt a standardised approach or an external ratings approach, the latter permitting banks to calculate capital charges based on broad groupings of the external ratings assigned by a recognized credit rating agency (Pasiouras et al. 2007). However, as the authors note, the disadvantage of the IRB approach is the extremely increased complexity regarding the technical aspects of the modelling. As such, similarly to Doumpos and Figueira (2019), we will adopt an external benchmarking framework for financial credit rating models developed internally by a financial institution. Therefore, the objective of the subsequent part of the analysis stems away from trying to 'replicate' CRAs' models, but rather uses the output (i.e. assigned credit rating) as external benchmark (that is nonetheless not necessarily error-free) of the internal evaluation.

3.2 Sample

Our sample consists of 55 listed EU³ banks rated by Standard and Poor's (S&P). This is a balanced panel for the period 2012–2017 (330 bank-year observations), with all data (ratings and risk assessment criteria) obtained through S&P's Market Intelligence Platform (formerly known as 'SNL'). The above frame is a product solely of availability according to the following criteria: (1) at least 6 years of data,⁴ (2) having a *balanced panel*, i.e. operating banks with available data for the whole period to be chosen—avoiding banks sporadically appearing and disappearing from our data set—and (3) excluding *defunct* banks as well as those being *M&A targets* at the time period examined. Enlarging the time period to the pre 2012 period would cause our sample to fall exponentially, thus we limited our analysis to this particular sample. In what follows, we go through the sample, giving more insights about it by focusing on both the response variable, detailed in Sect. 3.2.1, and the risk assessment criteria detailed in Sect. 3.2.2.

3.2.1 Credit ratings

As mentioned in the beginning of Sect. 3.2, the banks' credit ratings, for which we have available data are assigned by Standard & Poor's. Table 1 provides a tabulation of the banks in our sample per country. This is a balanced panel, so these banks are available for the whole period examined (i.e. 2012–2017). As reported in Table 1, banks in our sample operate across 19 EU countries, with Italy, Germany and France jointly representing a third of our sample, and Netherlands, UK and Spain approximately another fifth.

Additionally, Table 2 provides a tabulation of our sample according to our response variable. We have followed the same categorisation of S&P's notch scale as in the study of

³ We particularly restrict our research sample to EU banks mainly for two reasons. First, by including more regions around the globe, we would introduce a great deal of heterogeneity in our sample, which further to financial fundamentals would certainly require country-specific factors to be included, ranging from economic to regulatory ones. Second, by looking at the available data we need for this study across the globe, the EU region offered the most available data on a single region basis.

⁴ Our intention for availability of at least 6 years of data arises from the fact that, as it will be detailed in Sect. 4, we use the proposed model in three consecutive years (2015–2017) making use of a three-year window setting to define the typical profiles of reference actions.

Table 1 Sample composition by EU country

Country	Banks in sample	Bank-year obs.
Austria	3	18
Belgium	3	18
Bulgaria	1	6
Cyprus	1	6
Czech Republic	2	12
Denmark	2	12
Finland	3	18
France	6	36
Germany	6	36
Hungary	1	6
Ireland	1	6
Italy	7	42
Luxembourg	1	6
Netherlands	4	24
Poland	1	6
Portugal	2	12
Spain	4	24
Sweden	3	18
UK	4	24
Total	55	330

Our sample consists a strongly balanced panel of 55 EU banks operating across 19 bloc countries for the time-period 2012–2017; a total of 330 bank-year observations

Doumpos and Figueira (2019). In particular, the first class (C_1) includes highly speculative and non-investment grade rated banks, followed by lower (C_2) and upper (C_3) medium-rated banks, closing the list with the highest category (C_4) including top-rated banks. A delineation of this table can be graphically seen in Fig. 1.

As one could infer from the above figure, the lion share of the sample—with approximately 77.3% of bank-year observations—appears to be in the two middle categories (C_2 : 30.9%; C_3 : 46.4%), whilst the remaining 22.7% is split into the lower (C_1 : 14.5%) and higher (C_4 : 8.2%) categories. The evolution of the sample can be better delineated across the time span of our sample (2012–2017) in Fig. 2. Over the period 2012–2014 most banks seemed to be intact as to their credit ratings, with only a couple changes. In particular, two banks were ‘downgraded’ to C_2 from C_3 in 2014, with the remaining ones’ ratings staying intact. The remaining time period (i.e. 2015–2017) is characterised by more apparent deviations in the rankings. In particular, several banks improved their position by moving to a subsequent category and, just for the period 2015–2016, a few banks being downgraded from C_3 to C_2 .

3.2.2 Risk assessment criteria

The Uniform Financial Rating System (*UFIRS*), known as the CAMEL rating framework was introduced in 1979 by US regulators. The main purpose of this framework is to assess a bank’s financial condition around the following key areas: Capital adequacy (C), Asset

Table 2 Ratings in sample per category/class

Class	Grade	S&P	Description	Cases (bank-year obs.)	
				Grade	Class
C4	16	AAA	Prime	0	27
	15	AA+	High grade	6	
	14	AA		0	
	13	AA–		21	
C3	12	A+	Upper-medium	18	153
	11	A		89	
	10	A–		46	
C2	9	BBB+	Lower-medium	29	102
	8	BBB		34	
	7	BBB–		39	
C1	6	BB+	Non-investment	10	48
	5	BB–		11	
	4	BB		10	
	3	B+	Highly speculative	9	
	2	B		4	
	1	B–		4	

This table portrays the S&P's credit ratings of banks in our sample, and the total number of cases (bank-year observations)

quality (A), Management (M), Earnings (E), Liquidity (L). More precisely, *Capital adequacy* is related to the amount and quality of the institution's capital. *Asset quality* refers to the levels of existing and potential credit risk related to the institution's loan and investment portfolio. *Management* is a measure, which is primarily qualitative by nature and is related to the effectiveness of internal controls and audit systems. Furthermore, the *Management* component reveals further information related to board of director and managerial ability of meeting their goals. In regard to the *Earnings* component, this rates the bank's earnings, both current and expected. Finally, *Liquidity* is related to the assessment of the bank's ability to honour its cash payments as they fall due.

The CAMEL framework is an integral part of the assessment process carried out by central banks and regulatory bodies. It consists an internal supervisory tool on which regulatory authorities rely on in order to examine an institution around the five key areas outlined earlier. For example, in the US the three main financial regulatory agencies, namely the Federal Reserve, the Federal Deposit Insurance Corporation and the Office of Comptroller of the Currency carry out “on-site” and “off-site” examinations on a timely basis in order to monitor the safety and soundness conditions of the banks they supervise. During these examinations, the regulatory bodies rely on the CAMEL framework in order to identify the banks' strengths and weaknesses. Moreover, the findings are often considered early warning signs for an institution's financial health.

The main output of this examination is a composite rating ranging from 1 to 5. Ratings between 1 and 2 relate to low weakness, which can be controlled by the bank's management or board of directors. Ratings ranging between 3 and 4 refer to moderate to severe weaknesses, which could potentially not be able to be addressed by the institution's management. Finally, a rating of 5 is linked to critical weaknesses and poor safety

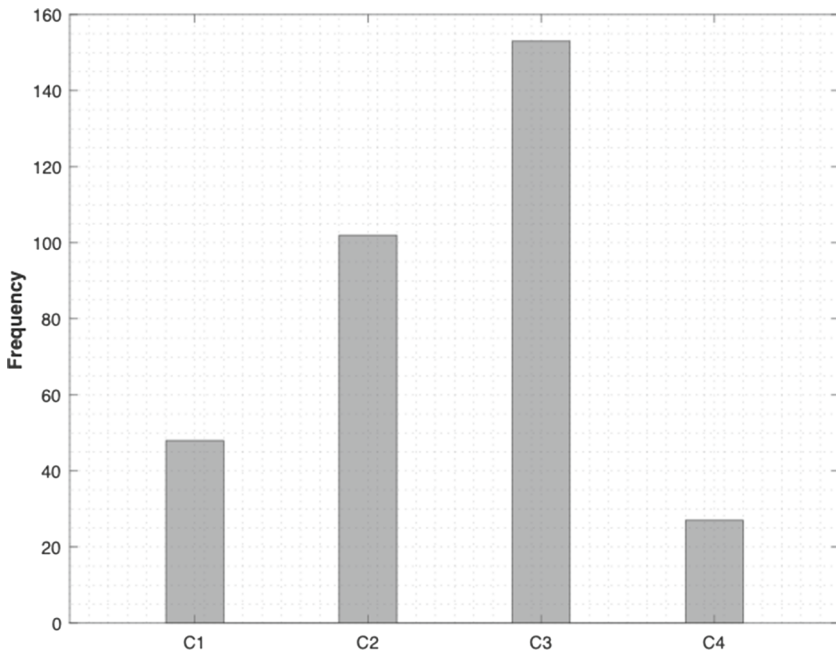
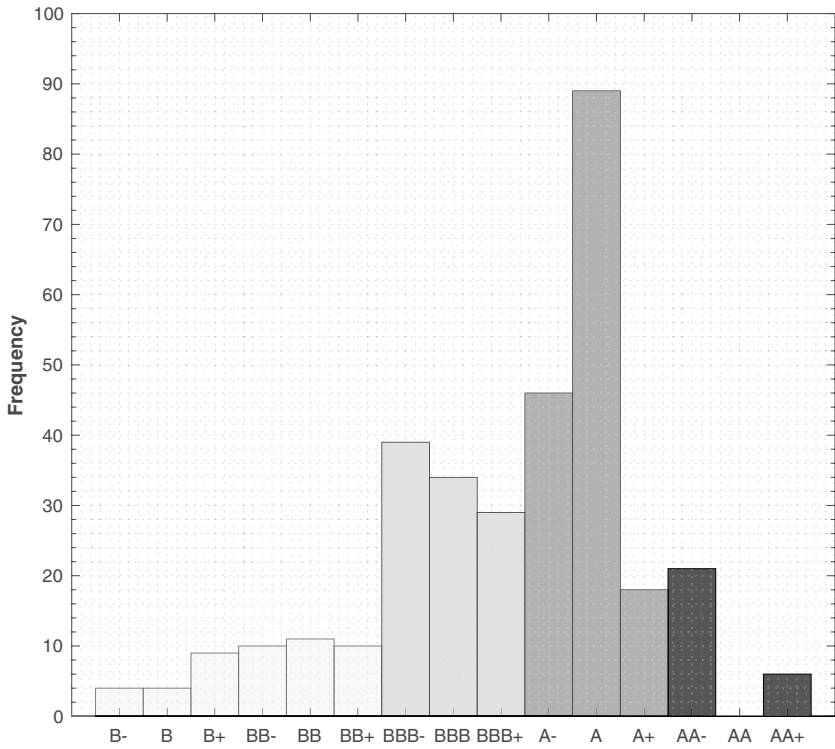
Fig. 1 Ratings composition by grade and class. This figure delineates Table 2. The upper part of the figure ▶ illustrates the composition of our sample according to the S&P's notch-scale ('grades' in Table 2), whilst the bottom part of the figure shows the sample's distribution according to the four classes (different gradient of grey) as these appear on the same Table

and soundness. It is important to note that the ratings are not publicly disclosed and only the senior management of a bank would be aware of the exact ratings. However, prior literature (e.g. Berger et al. 1998, 2001; Doumpos and Zopounidis 2010; Cole and White 2012) has suggested alternative ratios that could be used as proxies of each of the five components of this framework. We therefore rely on a set of financial characteristics that reflect a bank's financial condition in regard to its capital adequacy, asset quality, management, liquidity. Given that the list of candidate measures for each component could be non-exhaustive, we closely follow Doumpos and Zopounidis (2010) for the proxies taken into account. For each category, the authors list a variety of criteria. We choose those fulfilling the maximisation of our sample due to data availability. In particular, we measure *Capital Adequacy* with the ratio of Tier I and Tier II capital divided by the total risk-weighted assets, whereas as a proxy for *Asset Quality* we consider the risk-weighted assets ratio divided by the Total Assets of a firm. In order to capture the *Management* component, we make use of the staff costs as a fraction of the bank's total assets. The *Earnings* component is measured through the interest revenues to assets ratio. Finally, we account for the *Liquidity* component by including the ratio of cash & equivalent to total assets.

Turning to the descriptive statistics of the above-mentioned criteria, Table 3 shows their means in the whole sample tabulated according to the four credit rating classes. As it is expected, financial fundamentals improve in line with the credit ratings and even though differentiation between the lower categories might be small in some cases, the discriminating power of all criteria is confirmed at the 1% level of significance using the Kruskal–Wallis H test. A graphical complementary of Table 3 is available in Fig. 3a, which facilitates the reader's conception of the per class mean differences in each criterion, by illustrating the means along with a 95% confidence interval.

To give a sense of how this performance changes over time, Fig. 3b illustrates the average bank's performance in a three-year rolling window per class and per year for each criterion. Seemingly, top-rated banks significantly increased their performance in terms of capital adequacy ratio (just under 0.3 in 2015, just over 0.4 in 2016). They also cut staff costs (as % of their size) and lowered their risk-weighted assets. Non-investment banks on the other hand seem to have increased their staff costs and risk-weighted assets. Banks included in rating groups 2 and 3 (i.e. lower and upper-medium rated banks) did not significantly change their performance. While the latter remained more or less stable over time, the former marginally increased their capital adequacy ratios and liquidity and lowered their risk-weighted assets.

Table 4 presents the Pearson correlation coefficients. Evidently, a couple criteria are strongly correlated between them ('Interest Revenues (%TA)' with 'Risk-weighted Assets (%TA)', $\rho = 0.62$ and 'Staff costs (%TA)', $\rho = 0.64$). Nonetheless, it shouldn't pose a problem in the proposed model as it does not make inferences from the data, whilst the existence of a statistical association between the two does not necessarily mean causal relation, a case in which synergistic effects should be modelled and added to the model (see e.g. Corrente et al. 2017, for such a modelling in ELECTRE methods, and Arcidiano et al. 2018, for embedding this in PROMETHEE methods). In other words, we make



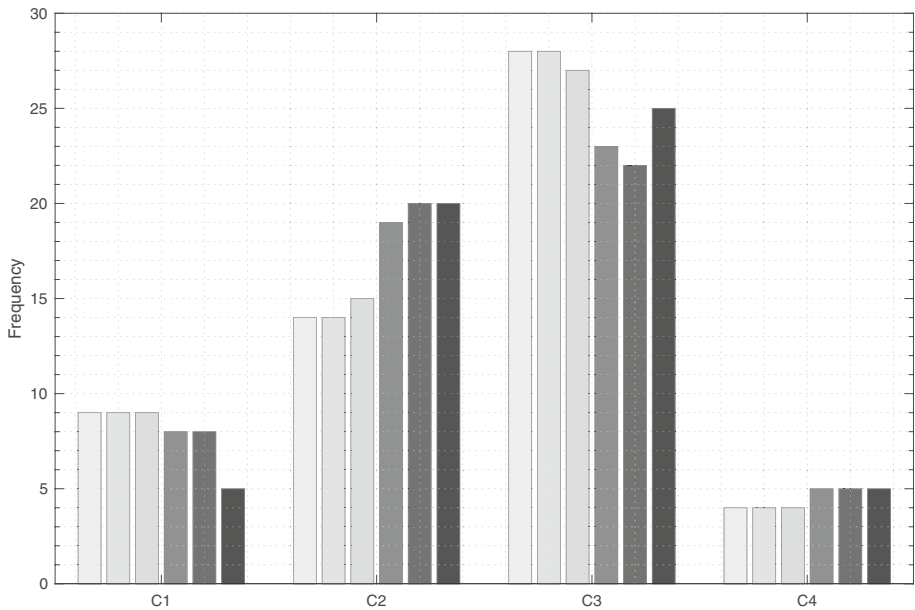


Fig. 2 Evolution of credit ratings across the time period examined. This figure delineates the evolution of banks' credit ratings across the 2012–2017 time period examined in our sample per each category (C1–C4) as these are disclosed in Table 2. Highlighted colours from light to dark grey illustrate the year, starting from 2012 and ending in 2017

Table 3 Means of the risk assessment criteria by credit rating category

Class	Capital adequacy ratio	Risk-weighted assets (%TA)	Staff costs (%TA)	Interest revenues (%TA)	Cash and equivalent (%TA)
C1	0.156	56.91	0.99	3.68	1.86
C2	0.167	45.09	0.81	2.52	1.05
C3	0.190	30.73	0.52	1.87	0.92
C4	0.346	22.51	0.26	1.43	0.05

This table shows the means of classes 1–4 according to the financial fundamentals chosen as risk-assessment criteria. For a graph, see Fig. 3

no assumption about such effects in the present analysis; yet, should the DM be aware of such effects, these should then be added and modelled in the framework accordingly. For instance, in the case of PROMETHEE methods (on which the sorting variant we hereby use is based on), Arcidiacono et al. (2018) offer a model based on which interactions could be modelled and visualised respectively.

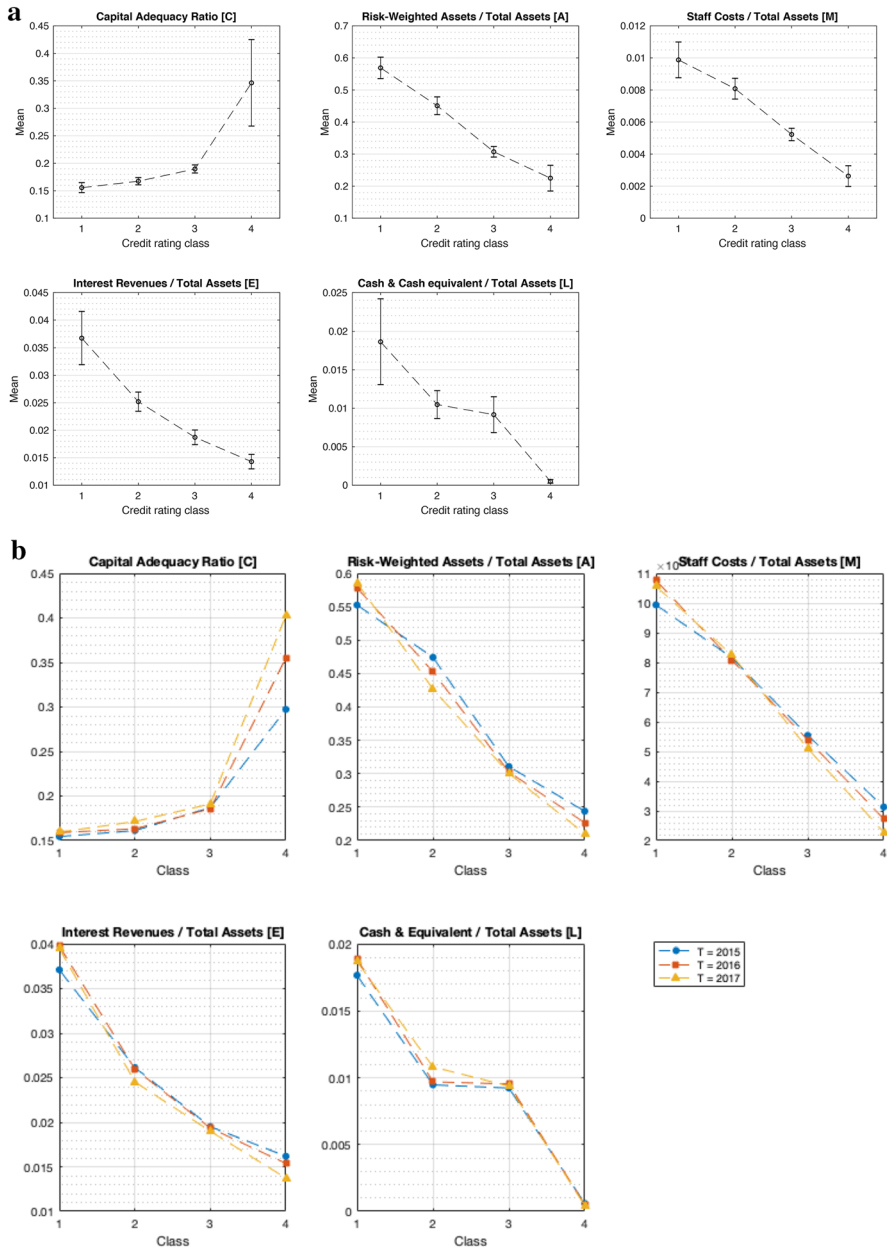


Fig. 3 **a** Means of the risk assessment criteria (w 95% CIs) by credit rating category. This figure delineates the means of the classes C1–C4 with their 95% confidence intervals for all five risk-assessment criteria. **b** Dynamic illustration of the typical (central) profiles of banks per class per year. This figure delineates the typical bank's performance per class and per year in the five risk-assessment criteria

Table 4 Pearson correlation

	Capital adequacy ratio	Risk-weighted assets (%TA)	Staff costs (%TA)	Interest revenues (%TA)
Risk-weighted assets (%TA)	-0.33	1		
Staff costs (%TA)	-0.23	0.56	1	
Interest revenues (%TA)	-0.21	0.62	0.64	1
Cash and equivalent (% TA)	-0.073	0.42	0.48	0.46

All relations are statistically significant at the 1% level

4 Empirical setting, results and discussion

This section presents (1) how the empirical setting around the sample described above was formulated, (2) the results obtained through the proposed method, as well as (3) their interpretation and a discussion based on them.

4.1 Modelling of parameters

Similar to the study of Doumpos and Figueira (2019), we use a rolling-window scheme involving three separate tests performed at three consecutive years, starting from 2015 and ending in 2017. In particular, for each year in this period, i.e. $T = 2015, 2016, 2017$, we employ data for that year and the previous two on a rolling-basis, in order to define the profiles of the reference alternatives on which in-sample banks are compared against in year T . Unlike the authors, we don't use a five-year rolling-window but rather a three-year one, as our data go back to 2012 and we intend to perform more than a single test. Regardless, using a similar setting to theirs does not significantly alter the results.

As described in Sect. 2, the SMAA-FFS approach requires a DM (e.g. hereby a credit analyst) to declare a set of parameters. These are the *limiting/central* profiles of each category, as well as the *weights* and *indifference/preference* thresholds for each criterion. In this analysis, on the basis of a lack of DM to give any kind of preference information (from the type of a probabilistic distribution, to assurance regions or any other constraints that could be used in the simulation environment), we uniformly simulate the required parameters. We give further information for each set of parameters below.

Starting with the profiles of each category for the year T , we first obtain from the rolling-window (historical) data for each criterion $j \in J$ the four means $(\mu_j^1, \mu_j^2, \mu_j^3, \mu_j^4)$ corresponding to the four true classes (C_1, \dots, C_4) , along with the minimum (x_j^{min}) and maximum (x_j^{max}) values observed for every criterion in this period. Then, the performance of a central profile in category $C_k, k = 1, \dots, 4$ on a risk assessment criterion $j \in J$ is assumed to be uniformly distributed in the following ranges:

$$\left\{ \begin{array}{ll} \left[x_j^{\min}, \frac{(\mu_j^1 + \mu_j^2)}{2} \right], & \text{if } k = 1 \\ \left[\frac{(\mu_j^{k-1} + \mu_j^k)}{2}, \frac{(\mu_j^{k+1} + \mu_j^k)}{2} \right], & \text{if } k = 2, 3 \\ \left[\frac{(\mu_j^3 + \mu_j^4)}{2}, x_j^{\max} \right], & \text{if } k = 4 \end{array} \right.$$

As also mentioned in Sect. 2, we make use of *central* instead of *limiting* profiles, as we believe that a reference point ‘behaving’ in a way that is described above would reasonably be a ‘typical’ representation of what a bank operating in a category $C_k, k = 1, \dots, 4$ could potentially look like.

Turning to the *indifference* (q_j) and *preference* (p_j) thresholds, these were not simulated in this analysis for reasons of simplicity. The reason being simulating such parameters would unnecessarily complicate the analysis, which is something found also in other similar studies (Angilella and Mazzù 2015; Doumpos and Figueira 2019, not simulating similar ELECTRE-based parameters). For this purpose, we set for each criterion j the indifference thresholds to 0 and the preference thresholds to the maximum attainable difference between alternatives $a_i \in A$ on criterion j . The drawback associated with this modelling specification though is that, in the virtue of simplicity, a fixed threshold like this means that the evaluation could be affected by outliers.

As far as the weighting of criteria is concerned, interaction with a DM could give us information about any type of constraints to be placed on the simulation process, e.g. inequality constraints (i.e. criterion j is more important than criterion j'), lower and upper bounds of criteria importance, or even customising the probabilities of random draws to a certain probabilistic distribution that is in line with what the DM(s) thinks feasible. Understandably, in the absence of any such information, which is the case in this illustrative analysis, we uniformly simulate the weights completely at random.⁵ Of course, this implies that the means of the criteria weights in this setting are approximately 20% for all criteria $j \in J$.

4.2 Predictive performance and analytics

In this subsection we describe a few ideas which could be -from an analytics perspective- value-adding being included in this kind of analysis. The first relates to the overall benchmarking against externally assigned credit ratings. Although the modelling framework presented so far is clearly not aiming to be in line with other predicting models used in the literature; benchmarking the internally computed ratings against the external ones is in line with optimising and revising the internal procedure. In that regard, the accuracy of a classification problem can be easily described using a confusion matrix. This is typically a $k \times k$

⁵ For an analytical guidance on this procedure, see point no.1 of Appendix I in the study of Doumpos et al. (2016). We hereby follow the same procedure in applying an infinitesimal constraint ε binding the lower feasible weight vector of each criterion j ($w_j \geq \varepsilon$) being simulated to avoid extreme scenarios in which a dimension is given a zero weight, thus implicitly excluding it from the analysis. In this case, we picked an ε value of 1%.

matrix, where each element c_{lm} in row $l = 1, \dots, k$ and column $m = 1, \dots, k$ describes the percentage of sample of true class m classified as class l . Put simply, the diagonal shows what percentage of expected classifications meets the actual ones, whilst the upper and lower triangles of the matrix showing respective deviations from t actual class. The overall accuracy (OA) (%) can be simply computed as follows:

$$OA = 100 \frac{\sum_{l=1}^k c_{ll}}{\sum_{l=1}^k \sum_{m=1}^k c_{lm}}.$$

This type of performance metric benchmarks the expected class (final classification as described in Sect. 2.3)—which takes into account all simulation outcomes into account via the use of one expected class- against the true class, as this is given by the CRA. Although this metric is interesting in itself and compares the overall expected classification against the true, externally assigned one; interesting insights could be obtained by looking at the overall accuracy (OA) of each classification taking into account each set of preferences expressed in the judgmental analysis simulation. This is an important metric for an analyst to consider, as one could explore the distribution of the achieved OAs according to the different preferences, and see whether any ‘extreme’ scenarios are included in this modelling framework, or isolate set(s) of preferences according to which this metric surpasses a given threshold (i.e. OA of at least 50%), or simply choose these preferences that maximise it overall. This could be obtained by simply computing an OA for each simulation, using instead of an overall expected classification, the classification that is achieved using the set of preferences in each simulation. This would create a $1 \times s$ vector, s being the number of simulations (hereby 10,000). By looking at the distribution of this vector, the analyst could consider the whole set of preferences, or -if a framework that is more in line with a predictive modelling is desired- discard those preferences that fall below a given threshold.

Complementing the OA metric, some analytics of interest to the credit analysts would be the following. First and foremost, we have talked about consolidating the category acceptability indices into a single, expected, classification through the intuition of a more ‘holistic’ aggregator. This gives the ordinal information necessary to conduct a standard classification model, and subsequently use it to obtain overall benchmarking metrics such as the one of a single overall OA mentioned above. Nonetheless, the information obtained from the category acceptability indices could simply be used to provide cardinal information on an important and interesting feature inherent in their use for credit rating analysis; that is their probabilistic indicators of rating membership according to the analysts’ perceptions taken into account in the judgmental analysis. In particular, these indicators illustrate for each bank the probability that it attains a given credit rating in the space of weight preferences taken into account. Such cardinal information complements the ordinal nature of a crisp class assignment and can provide interesting insights on the exact position of an alternative within a given class (Ishizaka et al. 2019).

Additionally, the typical shares of parameters classifying a bank in a particular category could be computed. For instance, as described in Sect. 2, central weight vectors show the typical analyst’s preferences on weights to classify a bank at a given category. Moreover, if one wishes to transform them into cumulative probabilistic indicators, concentration polarization indices (see e.g. Greco et al. 2018) could be used instead that measure e.g. the probability that a bank is classified in the bottom/top x categories, where x can be declared by the DM according to the notch-scale utilised in a classification setting (e.g. could be the bottom two, three, five etc. categories from the 22-notch scale of S&P).

Moreover, interesting insights could be obtained through visual aids, such as the (SMAA-)GAIA approach that could offer two types of information. First, looking at the GAIA plane, the credit analyst could observe the performance of a bank of interest relative to the criteria in which it is evaluated. For instance, if a bank is near a criterion's line on the plane and towards the same direction then this implies that it performs well on this criterion. On the contrary, if it is placed the opposite way, then it performs poorly on it. On the same plane, the credit analyst could also observe other banks behaving in a similar manner and, subsequently, find benchmark peers to contrast and compare against. The second type of analytics potentially of interest to the credit analyst relates to the stochastic nature of the analysis carried out so far and thus the SMAA version of GAIA (Arcidiacono et al., 2018) discussed in Sect. 2. In particular, the space of parameters (weights) could be plotted on the plane highlighting any kind of ordinal information which is greatly in line with a credit rating evaluation exercise to provide more insights on areas of improvement. For instance, the space of preferences could be highlighted on the GAIA plane exhibiting the classification ($C_k, k = 1, \dots, 4$) that a bank attains using those preferences.⁶ Alternatively, the credit analyst could simply visualise the space of parameters classifying the bank under evaluation in the worst (default) category (or a combination of relatively 'bad' categories). This could highlight areas of improvement for a given bank.

At this point, let us note something important about the last observation regarding the use of the SMAA-GAIA plane for analytics purposes. One should not confuse the benchmarking of a bank of interest with a reference profile as a prediction of external ratings. Put simply, by illustrating the space of parameters on the SMAA-GAIA plane, the credit analyst could have a direct evaluation with the central profile(s) set in the internal evaluation. That said, the overall prediction accuracy benchmark described in the beginning of this subsection illustrates the accuracy between our expected classification (i.e. taking into account all possible scenarios considered in the evaluation) and the true (externally assigned) ratings. This implies that by looking at *specific* benchmarks on the SMAA-GAIA plane does not necessarily mean that this prediction accuracy still holds. Rather, this type of evaluation is to be in an analyst's toolbox to compare how a bank of interest could perform against a central profile (hereby assumed to be that of the typical representative point from historical data) if certain parameters (e.g. weights) were such, or for reasons of sensitivity analysis.⁷

4.3 Results

In this section, we present the results obtained carrying out the analysis as described in Sect. 4.1 and discuss them based on the two types of metrics described in Sect. 4.2. When

⁶ For reasons of simplicity, the SMAA-GAIA plane illustrated in this analysis takes into account only one source of uncertainty: *weights*, holding the central profiles at their means. This permits direct and meaningful comparison among different set of weights that would otherwise be more complicated should central profiles were also changing.

⁷ Although it extends beyond the scope of this analysis, an interesting feature portraying that exact question (i.e. for every single set of preferences portrayed onto the plane, what is the overall prediction accuracy?) could be embedded onto the SMAA-GAIA plane. In particular, as each class is assigned a given colour in the RGB gamut, the intensity (transparency) of that colour could portray the accuracy of the prediction against external ratings. Nonetheless, let us not forget that the whole process of internal evaluation is not simply a matter of accurate external predictions, but also to highlight weaknesses and strengths of a bank itself, a scope of which the SMAA-GAIA plane adequately fulfils.

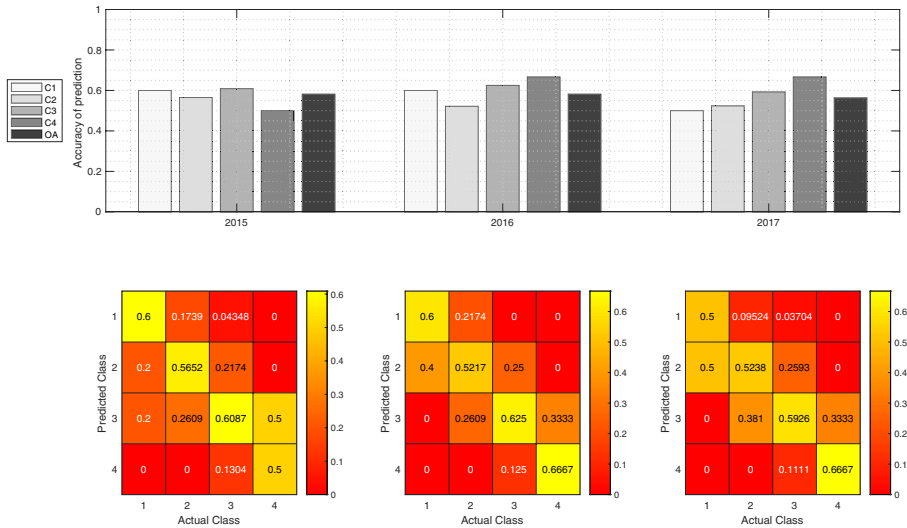


Fig. 4 Per class and overall accuracy of prediction. This figure delineates the overall accuracy of prediction (OA) of the expected classification (as described in Sect. 2.3). The lower part of the figure shows the confusion matrix for every year T that we put our model to use, whilst the upper part of the figure shows the diagonal of the confusion matrix for the same years and the overall accuracy of prediction (OA) for that year (see legend on the left part of the figure)

it comes to the latter, we give a simple illustration of how a credit analyst of a random bank in our sample could make insights based on this output. The illustrative example regards Cr dit Agricole, a French bank that in 2018 overthrew its US counterpart ‘Wells Fargo & Co.’ to take the No.10 spot in the world’s largest banks by total assets (Mehmood and Chaudhry 2018).

Starting with the accuracy metric, Fig. 4 delineates the overall accuracy (OA using the expected classification described at the end of Sect. 2.3) for each year we tested, $T=2015$, 2016, 2017. In particular, the confusion matrices are given in the bottom part of the figure, whilst their diagonals and OA metric are given in the upper part of the figure. As it is apparent from the figure, in all 3 years in which we tested this framework, we observed that -aside from a chance equal to a fifth classifying a C1-rated bank in C3 in year 2015- no bank that has been externally assigned to a C1-group is classified in C3 or C4 for years 2016 and 2017. On a similar note, no bank externally classified in C4 is found in our analysis to be classified in the bottom two categories. When it comes to the middle categories (i.e. C2 and C3), they overlap with modest probabilities into being classified between C1–C3 and C2–C4 respectively. Nonetheless, they are always classified into their classes with at least 52% probability. The overall accuracy (OA using the expected classification) of our prediction is just under 60% for years 2015 and 2016 (58.18% and 58.2% respectively) and falls by a slight amount into 56.4% in year 2017.

In order to get more insights on what happens in the simulation environment, the distribution of OAs is illustrated in Fig. 5a for the three-year period in which we run the analysis. These OAs are essentially computed using the formula given in Sect. 2.3, although instead of one final expected classification, the classification of each simulation ($s = 1, \dots, 10000$) was saved and benchmarked against the external, true classification of S&P. Figure 5a shows that in most simulations the predictive accuracy (against the externally assigned

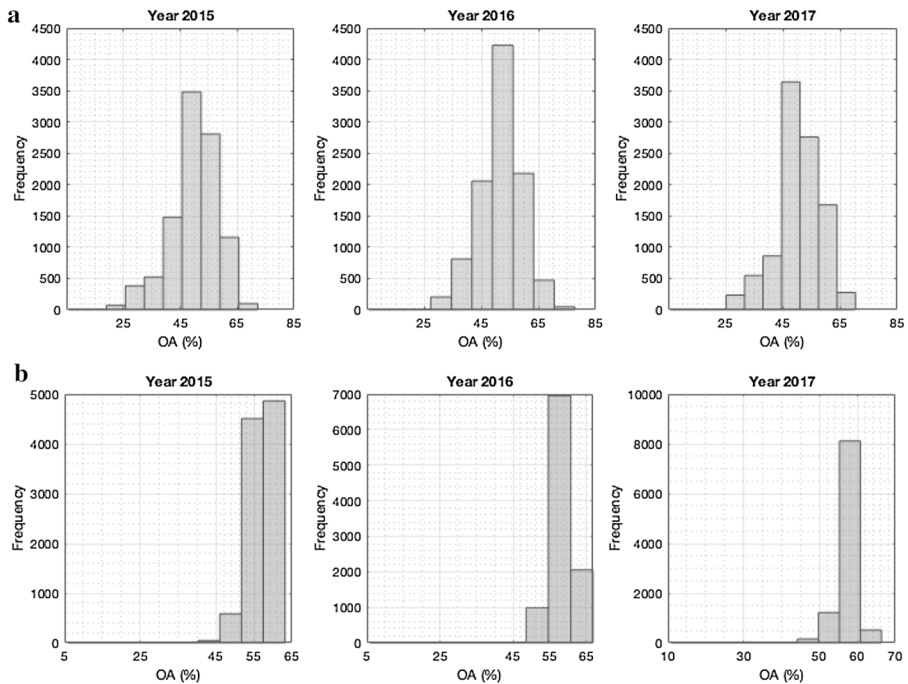


Fig. 5 **a** Distribution of OA in the simulations (unrestricted space W). This figure delineates how the in-simulation OAs (Sect. 4.2) are distributed in each year for which SMAA-FFS is employed. Results regard the unrestricted space W , i.e. weights follow a random uniform draw. **b** Distribution of OA in the simulations (restricted space W). This figure delineates how the in-simulation OAs (Sect. 4.2) are distributed in each year for which SMAA-FFS is employed. Results regard the restricted space W satisfying the following linear inequality: $w_A \geq w_C > w_E, w_M, w_L$

Table 5 Comparison with other classification models

Method	$T=2015$ (%)	$T=2016$ (%)	$T=2017$ (%)
Max SMAA-FFS	67.27	72.73	65.45
Linear Discriminant (LinDis)	60.00	64.30	63.60
UTADIS	69.30	69.10	65.50

This table shows the in-sample classification accuracy (OA) of three classification models. For the convenience of comparability, the best OA was chosen for SMAA-FFS, whilst Linear Discriminant (LinDis) and UTADIS were implemented without over-fit protection

ratings by the CRA) is beyond the 45 to 50% mark. In particular, it ranges between just under 18% and 67.3% for $T=2015$, just over 27% and 72.7% in $T=2016$ and just over 25% and 65.5% in $T=2017$.

As mentioned in Sect. 2.3, the space of weights W could be unrestricted (unconditional random draw of weights) or restricted (e.g. conditional on the fulfilment of some inequalities). If one assumes that in the context of European banks -in which our case study falls in- criteria such as ‘asset quality’ or ‘capital adequacy’ should be more important than the remaining ones, then the space of weights could be transformed accordingly through

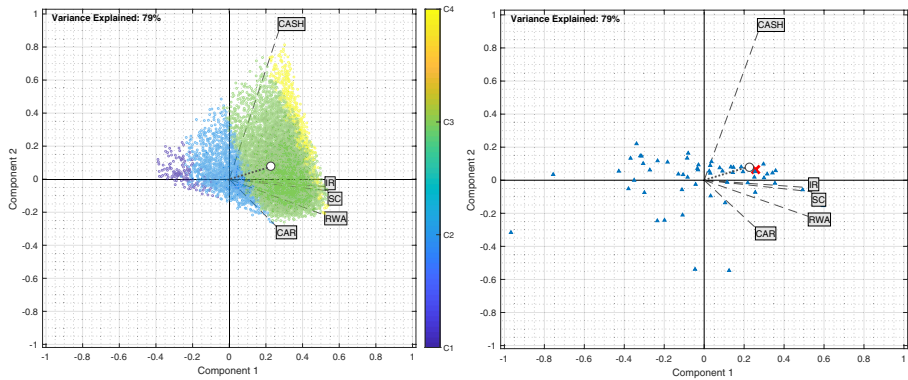


Fig. 6 SMAA-GAIA as a visual tool in an analyst's toolbox. This figure delineates the (SMAA-)GAIA plane in year 2017. Figure to the right shows the GAIA plane, in which the five risk assessment criteria ('CASH', 'IR', 'SC', 'RWA', and 'CAR') are drawn, alongside the vector of (equal) weights. Triangles illustrate the banks in our sample, with Crédit Agricole plotted with a red 'X'. The overall explained variance of the two components is 79%. On the left, the SMAA-GAIA plane for Crédit Agricole is displayed, again, along with the vector of equal weights to give a sense of space in the plane. Four different colours denote the space of parameters (weights) for which Crédit Agricole is classified in C1, C2, C3 and C4. This information is useful to the credit analyst, who can see under which linear combination of preferences Crédit Agricole is classified into a lower category than that desirable and vice versa and suggest immediate changes (e.g. focus on improving a certain dimension etc.). Another use relates to the sensitivity analysis that could be performed if the DM (here credit analyst) could give a single weight vector (suppose equal weights, like the one illustrated in the figure), and visualise how robust that evaluation is. For instance, if the analyst had favoured all dimensions the same (equal weights), then, seemingly, Crédit Agricole is fairly robust if changes relating to the preference of weights happen, given that the space that the bank is still classified as C3 is more than two-thirds of the overall feasible space of preferences

accommodating an inequality of the form e.g. $w_A \geq w_C > w_E, w_M, w_L$. This slightly increases the OA of our expected prediction by 0.4% in $T = 2015$, 0.2% in $T = 2016$ and 0.1% in $T = 2017$, though the main benefit of this constrained space is not the overall increased prediction, but a more 'stable' distribution as seen in Fig. 5b showing the distribution of OA in each simulation using this restrictive inequality.

Going back to the baseline model (i.e. unrestricted space W), our best in-sample prediction (taking into account the individual OAs in the simulation environment, i.e. Figure 5a) yields an accuracy of 72.7% in year 2016. If one wishes to see which weights contributed to this predictive performance, the central weight vectors can be computed (as described in Sect. 2.3), which are essentially the typical (average) preferences assigning an OA score of 72.7%. In our case, out of the 10,000 simulations, there were two particular weight vectors giving a score of $OA = 72.7\%$. Thus, the average of these preferences was computed, which formed the vector $w = [0.0754, 0.3184, 0.2606, 0.3073, 0.0383]$. This means that the most important criterion for in-sample prediction accuracy in 2016 was asset quality (31.84%), followed by earnings (30.73%), management (26.06%), capital adequacy (7.54%) and liquidity (3.83%). Very similar weight vectors were used in years 2015 and 2017 too.

A natural question that may arise is how other models specifically designed to maximise the prediction accuracy perform in this setting. We have compared our model to an MCDA disaggregation, such as UTADIS (Zopounidis and Doumpos 1999, 2002), as well as Linear Discriminant Analysis (LinDis). In order to make approaches as comparable as possible, we used the maximum attainable prediction of OA for the SMAA-FFS model, and we used

UTADIS and LinDis without over-fit protection, i.e. their results regard in-sample prediction. Results in Table 5 reveal that both UTADIS and SMAA-FFS seem to perform better than LinDis for all three years in our sample; yet, UTADIS outperforms SMAA-FFS in both years 2015 and 2017. Regardless, we should note again that, first the aim of this approach lies beyond that of a predictive model that is based on statistical approaches or linear programming, and second, it is mainly the product of banks' risk assessment criteria not hugely overlapping across categories that these results can be attributed to, i.e. the average performance of a bank on the CAMEL framework criteria is increasing with the credit ratings assigned with no huge deviations from these means as it is apparent in Fig. 3a.

Figure 6 illustrates the (SMAA-)GAIA graphical illustration described in Sect. 4.2, using the example of *Crédit Agricole*. On the right part of the figure we can see the GAIA plane of the evaluation for the year 2017. To begin with, plotted on the plane are the five risk assessment criteria representing the respective proxies of the CAMEL framework. The length of the criteria lines shows the discriminating power they have. Seemingly the proxy for Liquidity ('CASH') seems to be slightly more discriminating when it comes to the results, with the remaining four criteria having approximately the same discriminating power. Moreover, lines of criteria 'IR', 'SC' and 'RWA' are plotted almost next to each other, which essentially means that a bank that performs well on one is expected to equally perform well on the remaining two, something that is expected given their moderately strong correlation approximately around the 0.6 range. 'CASH' seems to be orthogonal to the remaining criteria, which indicates no particular meaningful relation there, whilst 'CAR' is more independent than the three correlated risk assessment criteria. Banks plotted on the bottom left part of quadrant 3 seem to be dominated as to the liquidity ('CASH') criterion, whereas banks on quadrant 4 are generally performing poorly with respect to the other 4 criteria ('CAR', 'RWA', 'SC', 'IR') that are plotted in the opposite way. Taking the example of *Crédit Agricole*, it is among a cluster of banks performing particularly well as to the Earnings and Management proxies of the framework in question.

Turning to the left plot of Fig. 6, it delineates a credit analyst's desire to see more insights as to a particular bank of interest. Supposedly, the analyst is interested again in *Crédit Agricole*'s evaluation. Having given a range of potential preferences for this analysis (here we assume that the whole space of weights could be deemed feasible by the analyst), what is apparent is for each (weight) vector of preferences, what is the classification achieved for the bank in question. In particular, every single plotted point on the plane characterises a potential vector of preferences that is coloured in such a way to visualise what is the classification outcome of *Crédit Agricole* respectively. Technically speaking, every such point is a different 'decision stick' (see Mareschal and Brans 1988) on the GAIA plane. Plotted alongside the rest is a vector of equal weights (white circle with dotted line) that gives a sense of direction on where, in this two-dimensional plane, a vector of equal preferences would fall on. This figure can also act as a tool of sensitivity analysis, illustrating the size of the space by which, changing the preferences of a DM (credit analyst doing the internal evaluating) holds the classification intact. For example, taking into account all potential 10,000 weight vectors, *Crédit Agricole* seems to be fairly robust, with more than two-thirds of the space of preferences covering C3 ('green colour'), the actual class that it also belongs to according to the external ratings.

Additionally, given the nature of the panel data in our sample, one could be interested in tracking how the performance of banks changes over time. One way that this can be achieved with is by looking at the average net flows of the classes across the time period examined. Figure 7 shows this trend. Seemingly, top-rated banks significantly stepped up their performance in year 2016 and their growth remained stagnant the following year

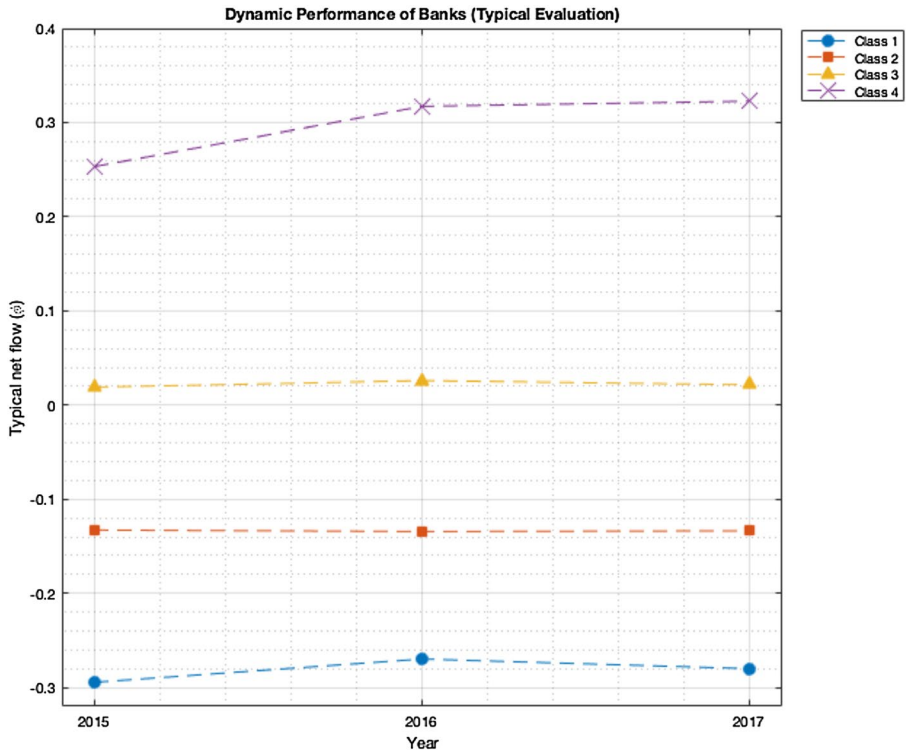


Fig. 7 PROMETHEE evaluation of typical banks per class and per year. This figure delineates the how the average bank in each class is evaluated (average net flow ϕ , see Sect. 2.3, taking into account all sampled weights) and how this changes dynamically across the time period 2015–2017 examined

(presenting a marginally noticeable increase). Class 3 banks gained marginal traction in year 2016 but lost most of it the following year. Similar patterns appear for non-investment banks, while lower-medium rated banks' performance seemed to be stable the whole period 2015–2017.

A second way to track performance is to look at the probabilistic classification of each bank, or groups of banks overall. In particular, in Sect. 4.2 we discussed how the probabilistic classification outcomes within the SMAA-FFS framework could disseminate cardinal information that shows the degree of membership to each pre-defined rating group. In this illustrative example of Crédit Agricole for instance, in year 2017 it received only a 1.46% chance to be classified in the worst class (C1), a 15.98% chance to be in the second class (C2) and a 77.43% chance to be classified in the category where it rightfully belongs (according to external ratings), whilst for a more modest 5.13% chance it was classified as a high-grade/prime-rated bank. The previous year, the same bank achieved a group membership of 2.29%, 15.72%, 69.19% and 12.80% respectively, which shows that the bank's performance slightly deteriorated against the typical bank top-rated bank (C4), although slightly improved against the typical worst-rated bank (C1) and against the typical upper-medium rated bank (C3). The full set of results for all 55 banks and for each year are not reported to conserve space, but are provided in the on-line supplementary appendix for the interested reader.

Finally, as noted in Sect. 4.2 there are two things to be aware of regarding these memberships. First and foremost, these are not a guarantee of external ratings, but rather part of the internal evaluation process. That said, they only reflect the opinions of the analyst(s) involved in this process (through the sets of preferences taken into account in the simulation). Second, in a more high-dimensional classification setting involving shorter notchescales (e.g. 8-10 categories or more), one could use polarization indicators that are actually the aggregate probability of being classified, e.g. in the bottom five categories etc.

5 Conclusive remarks

In this study, we have introduced an application of the recently-introduced SMAA-Fuzzy-FlowSort approach to the case of modelling bank credit ratings. The stochastic nature of the approach welcomes handling of imprecisions and uncertainty around several parameters of the modelling phase, which is naturally in line with a decision-making exercise of this type. Moreover, the category acceptability indices produced as an output of this approach complement the ordinal nature of a judgmental rating procedure with cardinal information that shows the probability of a bank to be classified into a given category. Put simply, they encapsulate a type of uncertainty analysis to illustrate banks' probabilistic memberships, but also an expected overall classification. Combining the above output with the SMAA variant of GAIA offers a credit analyst with a visual of a bank's judgmental analysis, complementing this modelling framework with a more a holistic multicriteria decision support tool that enables a rich inferential procedure to be conducted. To illustrate the assets of this framework, we have provided a simple case study involving the evaluation of the credit risk of 55 EU banks according to a handful financial fundamentals.

As far as the future avenues of research are concerned, from a methodological viewpoint, embedding features permitting the handling of interactions among criteria in conjunction with a framework that models a hierarchical structure (Arcidiacono et al. 2018; Corrente et al. 2017) would certainly be of interest, whilst it would provide a more complete model that does not make assumptions like the ones we did in this study assuming no externalities between attributes. From an empirical perspective, even though we have hereby used a basic CAMEL-based model, including more qualitative factors aside from simply bank fundamentals would certainly be interesting and enriching, as well as closer to the CRA's selected factors according to their handbook (see e.g. S&P 2012a, b). The inclusion of further criteria could potentially increase the accuracy against externally assigned ratings, whilst embedding a hierarchical structure modelling such as the MCHP (Corrente et al. 2012) could give the credit analyst even more insights about how each dimension performs and changes over time (see, Corrente et al. 2017, for an example of a classification making use of MCHP in a case study involving CAMEL ratings). In this proposal we kept both the number of criteria and assumptions about externalities as simple as possible to showcase the importance and capabilities of the presented tools. Regardless, the aforementioned extensions could well-benefit the analyst in real world scenarios that involve more complex dynamics.

Acknowledgements We would like to thank two anonymous reviewers for their comments and enriching ideas that greatly helped us to improve this manuscript. All errors remain our own.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allen, R., Athanassopoulos, A., Dyson, R. G., & Thanassoulis, E. (1997). Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of Operations Research*, 73, 13–34.
- Almeida-Dias, J., Figueira, J. R., & Roy, B. (2012). A multiple criteria sorting method where each category is characterized by several reference actions: The Electre Tri-nC method. *European Journal of Operational Research*, 217(3), 567–579.
- Angilella, S., & Mazzù, S. (2015). The financing of innovative SMEs: A multicriteria credit rating model. *European Journal of Operational Research*, 244(2), 540–554.
- Arcidiacono, S. G., Corrente, S., & Greco, S. (2018). GAIA-SMAA-PROMETHEE for a hierarchy of interacting criteria. *European Journal of Operational Research*, 270(2), 606–624.
- Basel Committee on Banking Supervision (2011). Basel III: A Global regulatory framework for more resilient banks and banking systems. Revised version June 2011.
- Behzadian, M., Kazemzadeh, R. B., Albadvi, A., & Aghdasi, M. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, 200(1), 198–215.
- Belton, V., & Stewart, T. (2002). *Multiple criteria decision analysis: An integrated approach*. Dordrecht: Springer.
- Berger, A. N., Davies, S. M., & Flannery, M. J. (1998). Comparing market and supervisory assessments of bank performance: Who knows what when? FEDS Paper No.98-32.
- Berger, A. N., Kyle, M. K., & Scalise, J. M. (2001). Did US bank supervisors get tougher during the credit crunch? Did they get easier during the banking boom? Did it matter to bank lending? In F. S. Mishkin (Ed.), *Prudential supervision: What works and what doesn't* (pp. 301–356). Chicago: University of Chicago Press.
- Brans, J. P., & De Smet, Y. (2016). PROMETHEE methods. In S. Greco, M. Ehrgott, & J. R. Figueira (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 187–229). Berlin: Springer.
- Brans, J. P., & Mareschal, B. (1995). The PROMETHEE VI procedure: how to differentiate hard from soft multicriteria problems. *Journal of Decision Systems*, 4(3), 213–223.
- Brans, J. P., & Vincke, P. (1985). Note—a preference ranking organisation method: (the PROMETHEE method for multiple criteria decision-making). *Management Science*, 31(6), 647–656.
- Brans, J.-P., Vincke, P., & Mareschal, B. (1986). How to select and how to rank projects: The PROMETHEE method. *European Journal of Operational Research*, 24(2), 228–238.
- Campos, A. C. S. M., Mareschal, B., & de Almeida, A. T. (2015). Fuzzy FlowSort: An integration of the FlowSort method and fuzzy set theory for decision making on the basis of inaccurate quantitative data. *Information Sciences*, 293, 115–124.
- Capotorti, A., & Barbanera, E. (2012). Credit scoring analysis using a fuzzy probabilistic rough set model. *Computational Statistics & Data Analysis*, 56(4), 981–994.
- Cole, R. A., & White, L. J. (2012). Déjà vu all over again: The causes of US commercial bank failures this time around. *Journal of Financial Services Research*, 42(1–2), 5–29.
- Corrente, S., Doumpos, M., Greco, S., Słowiński, R., & Zopounidis, C. (2017). Multiple criteria hierarchy process for sorting problems based on ordinal regression with additive value functions. *Annals of Operations Research*, 251, 117–139.
- Corrente, S., Figueira, J. R., & Greco, S. (2014). The SMAA-PROMETHEE method. *European Journal of Operational Research*, 239(2), 514–522.
- Corrente, S., Greco, S., & Słowiński, R. (2012). Multiple criteria hierarchy process in robust ordinal regression. *Decision Support Systems*, 53(3), 660–674.
- Demircuc-Kunt, A., Detragiache, E., & Tresselt, T. (2006). Banking on the principles: compliance with Basel core principles and bank soundness. World Bank Policy Research Working Paper 3954.

- Doumpos, M., & Figueira, J. R. (2019). A multicriteria outranking approach for modeling corporate credit ratings: An application of the Electre Tri-nC method. *Omega*, 82, 166–180.
- Doumpos, M., Gaganis, C., & Pasiouras, F. (2016). Bank diversification and overall financial strength: International evidence. *Financial Markets, Institutions & Instruments*, 25(3), 169–213.
- Doumpos, M., & Pasiouras, F. (2005). Developing and testing models for replicating credit ratings: A multicriteria approach. *Computational Economics*, 25(4), 327–341.
- Doumpos, M., & Zopounidis, C. (2004). A multicriteria classification approach based on pairwise comparisons. *European Journal of Operational Research*, 158(2), 378–389.
- Doumpos, M., & Zopounidis, C. (2010). A multicriteria decision support system for bank rating. *Decision Support Systems*, 50(1), 55–63.
- Doumpos, M., & Zopounidis, C. (2014). *Multicriteria analysis in finance*. New York: Springer.
- Gaganis, Ch., Pasiouras, F., & Zopounidis, C. (2006). A multicriteria decision framework for measuring banks' soundness around the world. *Journal of Multi-Criteria Decision Analysis*, 14(1–3), 103–111.
- García, F., Giménez, V., & Guijarro, F. (2013). Credit risk management: A multicriteria approach to assess creditworthiness. *Mathematical and Computer Modelling*, 57(7–8), 2009–2015.
- Gavalas, D., & Syriopoulos, T. (2015). An integrated credit rating and loan quality model: application to bank shipping finance. *Maritime Policy & Management*, 42(6), 533–554.
- Greco, S., Ehrgott, M., & Figueira, J. (2016). *Multiple criteria decision analysis: State of the art surveys*. Berlin: Springer.
- Greco, S., Ishizaka, A., Matarazzo, B., & Torrisi, G. (2018). Stochastic multi-attribute acceptability analysis (SMAA): an application to the ranking of Italian regions. *Regional Studies*, 52(4), 585–600.
- Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). Sigma-Mu efficiency analysis: A methodology for evaluating units through composite indicators. *European Journal of Operational Research*, 278(3), 942–960.
- Grunert, J., Norden, L., & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance*, 29(2), 509–531.
- Ishizaka, A., Tasiou, M., & Martínez, L. (2019). Analytic hierarchy process-fuzzy sorting: An analytic hierarchy process-based method for fuzzy classification in sorting problems. *Journal of the Operational Research Society*. <https://doi.org/10.1080/01605682.2019.1595188>.
- Lahdelma, R., Hokkanen, J., & Salminen, P. (1998). SMAA-stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, 106(1), 137–143.
- Lahdelma, R., & Salminen, P. (2001). SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. *Operations Research*, 49(3), 444–454.
- Lahdelma, R., & Salminen, P. (2010). A method for ordinal classification in multicriteria decision making. In *Proceedings of the 10th IASTED international conference on artificial intelligence and applications* (Vol. 674, pp. 420–425).
- Langohr, H. M., & Langohr, P. T. (2010). *The rating agencies and their credit ratings: What they are, how they work and why they are relevant*. Hoboken: Wiley.
- Mareschal, B., & Brans, J. P. (1988). Geometrical representations for MCDA. *European Journal of Operational Research*, 34(1), 69–77.
- Mehmood, J. Z., & Chaudhry, S. (2018). The world's 100 largest banks. SNL Financial Extra, 6 April 2018. Retrieved March 24, 2018, from [https://platform.mi.spglobal.com/web/client?auth=inherit"V"news/article?id=44027195&cdid=A-44027195-11060](https://platform.mi.spglobal.com/web/client?auth=inherit).
- Nemery, P., & Lamboray, C. (2008). FlowSort: A flow-based sorting method with limiting or central profiles. *TOP*, 16, 90–113.
- Pasiouras, F., Gaganis, C., & Doumpos, M. (2007). A multicriteria discrimination approach for the credit rating of Asian banks. *Annals of Finance*, 3(3), 351–367.
- Pasiouras, F., Gaganis, C., & Zopounidis, C. (2006). The impact of bank regulations, supervision, market structure and bank characteristics on individual bank ratings: A cross-country analysis. *Review of Quantitative Finance and Accounting*, 27(4), 403–438.
- Pelissari, R., Oliveira, M. C., Amor, S. B., & Abackerli, A. J. (2019a). A new FlowSort-based method to deal with information imperfections in sorting decision-making problems. *European Journal of Operational Research*, 1(1), 235–246.
- Pelissari, R., Oliveira, M. C., Amor, S. B., Kandakoglu, A., & Helleno, A. L. (2019b). SMAA methods and their applications: A literature review and future research directions. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-019-03151-z>.
- Poon, W. P. H., & Firth, M. (2005). Are unsolicited credit ratings lower? International evidence from bank ratings. *Journal of Business Finance and Accounting*, 32(9&10), 1741–1771.

- Poon, W. P. H., Firth, M., & Fung, H.-G. (1999). A multivariate analysis of the determinants of Moody's bank financial strength ratings. *Journal of International Financial Markets, Institutions and Money*, 9, 267–283.
- Roy, B., & Bouyssou, D. (1993). *Aide multicritère à la décision: méthodes et cas* (p. 695). Paris: Economica.
- Standard and Poor's (S&P). (2012a). General criteria: Methodology: Management and governance credit factors for corporate entities and insurers, November 13.
- Standard and Poor's (S&P). (2012b). How we use management and governance credit factors, November 13.
- Tervonen, T., Lahdelma, R., Dias, J. A., Figueira, J., & Salminen, P. (2007). SMAA-TRI. In I. Linkov, G. A. Kiker, & R. J. Wenning (Eds.), *Environmental security in harbours and coastal areas* (pp. 217–231). Dordrecht: Springer.
- Treacy, W. F., & Carey, M. (2000). Credit risk rating systems at large US banks. *Journal of Banking & Finance*, 24(1–2), 167–201.
- White, L. J. (2010). Markets: The credit rating agencies. *Journal of Economic Perspectives*, 24(2), 211–226.
- Zopounidis, C., & Doumpos, M. (1999). Business failure prediction using the UTADIS multicriteria analysis method. *Journal of the Operational Research Society*, 50(11), 1138–1148.
- Zopounidis, C., & Doumpos, M. (2002). Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research*, 138(2), 229–246.
- Zopounidis, C., Galariotis, E., Doumpos, M., Sarri, S., & Andriosopoulos, K. (2015). Multiple criteria decision aiding for finance: An updated bibliographic survey. *European Journal of Operational Research*, 247(2), 339–348.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.